A case for corpus stylistics: Ian Fleming's Casino Royale

Mahlberg, Michaela; McIntyre, Dan

DOI:
10.1075/etc.4.2.03mah

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Mahlberg, M & McIntyre, D 2011, 'A case for corpus stylistics: Ian Fleming's Casino Royale', *English Text Construction*, vol. 4, no. 2, pp. 204-227. https://doi.org/10.1075/etc.4.2.03mah

A case for corpus stylistics

Ian Fleming's *Casino Royale*

Michaela Mahlberg and Dan McIntyre
University of Nottingham / University of Huddersfield

In this article we investigate keywords and key semantic domains in Fleming's *Casino Royale*. We identify groups of keywords that describe elements of the fictional world such as characters and settings as well as thematic signals. The keyword groups fall into two broad categories that are characterized as text-centred and reader-centred, with the latter providing particular clues for interpretation. We also compare the manually identified keyword groups with key semantic domains that are based on automatic semantic analysis. The comparison shows, for instance, how words that do not seem to fit a semantic domain can be seen as reader-centred keywords fulfilling specific textual functions. By linking our analysis to arguments in literary criticism, we show how quantitative and qualitative approaches can usefully complement one another.

1. Introduction: A corpus approach to popular fiction

In an influential article on semantic prosody and its potential for creating ironic effects in texts, Bill Louw (1993) criticises Donald C. Freeman's dismissal of the notion that it would ever be possible to ascertain frequency distributions for particular linguistic features of discourse (Freeman 1970: 3) and hence a norm against which stylistic deviation could be measured. Notwithstanding the fact that in 1970 computational methods of linguistic analysis were still in their infancy, Freeman's view is surprisingly uncompromising. He goes so far as to say that "even if they [frequency distributions] could be ascertained they would constitute no revealing insight into either natural language or style" (Freeman 1970: 3). In his article, Louw (1993) shows the imprudence of such a position by using large computer corpora to shed light on how irony is conveyed in texts through what he terms 'semantic prosodies'.[1] Louw's work rests on his being able to compare particular collocations against the norms apparent in the Bank of English. For instance, he shows how David Lodge's use of the phrase *bent on* in his novel *Small World* gives rise to an

English Text Construction 4:2 (2011), **204–227**. DOI 10.1075/etc.4.2.03mah
ISSN 1874–8767 / E-ISSN 1874–8775

ironic effect because *bent on* is usually used in reference to negatively-charged pursuits, whereas Lodge uses it in relation to the practice of self-improvement (see Louw 1993: 164–166), thus foregrounding the phrase. Louw's research provides convincing evidence of the potential for corpus linguistic methodologies to benefit stylistic analysis, consequently invalidating Freeman's now outdated claim. Indeed, in recent years the use of corpus techniques in stylistics has become more widespread (see, for example, Adolphs & Carter 2002, Hori 2004, Semino and Short 2004, Stubbs 2005, Culpeper 2009, McIntyre and Walker 2010, or for more general overviews see Biber 2011, Mahlberg forthcoming, McIntyre forthcoming) giving rise to a new sub-branch of stylistics increasingly known as 'corpus stylistics'. Our aim in this article is to contribute to this emerging field by demonstrating the interplay of qualitative and quantitative methods in a corpus stylistic analysis of Ian Fleming's novel *Casino Royale*. We are particularly interested in the question of how keywords can usefully be categorized for the analysis of literary texts.

Generally, popular fiction has not received the same degree of attention from literary scholars as have those works commonly classed as 'literature', despite the demonstration by stylisticians that 'literary language' cannot be clearly delineated from everyday language (see, for example, Carter and Nash 1990, Carter 2004). Those analyses of popular fiction that have been produced tend to approach the genre using a qualitative analytical methodology (e.g. Nash 1990, Montoro 2007). This is not to say that a qualitative approach is not valid (indeed, it should be a crucial part of any stylistic analysis), but simply that qualitative analyses might benefit from the added insights that a corpus approach can provide.

We decided to focus initially on Ian Fleming's James Bond novels, since this body of work has achieved a significant degree of popular success (i.e. large sales though little critical acclaim from the literary establishment). Additionally, the series of novels provides the opportunity to look at certain stylistic issues — for instance, character development — over a period of time. Fleming is also noted for his effusive writing style, often referred to by Bond fans as 'the Fleming sweep' and accounted for in part by his admission to writing the novels quickly in a relatively short period (usually around six weeks). In this article we focus on the first novel in the series, *Casino Royale* (Jonathan Cape, 1953), henceforth *CR*, and demonstrate how a corpus analysis can provide insights into characterization, the creation of particular stylistic effects and the construction of the fictional world of the text.

A major benefit of using corpus techniques to aid stylistic analysis is that this practice enables us to address what has long been an issue with the analysis of prose fiction. This is the problem of length and the fact that most prose texts are simply too long for the stylistician to deal with. Leech and Short (2007) point out that this is one of the main difficulties associated with the analysis of prose texts:

> […] the sheer bulk of prose writing is intimidating; […] In prose, the problem of how to select — what sample passages, what features to study — is more acute, and the incompleteness of even the most detailed analysis more apparent.
> (Leech and Short 2007: 2)

Using corpus linguistic methodologies can assist in determining the selection of examples to study, as we shall show, as well as providing a means of examining whole texts as opposed to small extracts. Of course, this is not to suggest that corpus techniques should replace manual analysis. A corpus analysis will still be incomplete, but it can help to cover additional aspects of the text in order to complement the manual analysis. It thus adds further techniques to the stylistician's 'toolkit'. The ability to look at whole texts also enables us to return to one of the earliest concerns of stylistics, which was the analysis of the style associated with individual authors and works. Although it is beyond the scope of the present article to deal with this entirely, our analysis here of one novel may be taken as a starting point for an analysis of Fleming's complete series of Bond books. The fact that stylisticians see an advantage in corpus stylistics for helping with the analysis of 'long' texts also highlights that corpus stylistics adds innovation to both literary stylistics and corpus linguistics. In corpus linguistics, a single novel would be regarded as a rather small sample of data that is unlikely to receive much attention as a text in own right, but is more likely to be included as part of a corpus.

2. Keywords and the stylistician's checklist

With the present analysis we aim to (i) illustrate how corpus methods can add to identifying lexis and associated text extracts for further qualitative analysis and (ii) explore how such a corpus approach can link in with concerns in literary stylistics and criticism. The tool we used is the WMatrix software package (Rayson 2008). The distinguishing feature of this corpus software is the automatic semantic analysis of texts. Options in WMatrix also include the generation of keywords and concordances. These techniques might be seen as constituting a checklist for corpus stylistic analysis in a similar way as Leech and Short (2007) suggest that lexis, grammar, figures of speech, cohesion, etc. serve as checklist items for qualitative stylistic analysis. It is important to emphasize, however, that simply generating keywords does not constitute an analysis. From the range of linguistic devices that stylisticians have at their disposal for the analysis of a literary text, a useful choice is usually motivated by the analyst's interpretation of the text. In this sense, keywords illustrate a different kind of approach. Any text will yield some set of keywords and the point of the automatic generation of keywords is that the computer generates them 'blindly' without seeing meaning relationships between them (cf.

Scott 2010: 45f.).The 'meaningfulness' of keywords is then revealed in the qualitative analysis.

Keywords are words that are significantly more frequent in a text or set of texts than in a reference corpus. Interpreted in stylistic terms, the reference corpus can be seen as providing a norm, so that keywords are words whose frequency deviates from that norm. Such deviations are of interest in stylistic analysis, where the concept of foregrounding is used to describe the effects that result from deviations from linguistic norms. Thus for corpus stylistic studies, corpora and the norms they provide can aid the identification of linguistic phenomena that point to stylistic features worthy of further analysis. For examples of keyword studies of literary texts see e.g. Culpeper (2009), Scott and Tribble (2006), O'Halloran (2007), Toolan (2009). While the analysis of keywords is becoming increasingly popular in corpus linguistics, keyness can still be seen as "a new territory" and "there are problems in identifying exactly what the keyness procedure comes up with and determining the boundaries of the research endeavour" (Scott 2010: 43). Words that are highlighted by the computer as key seem to broadly fall into three categories: names or proper nouns, words that indicate aboutness (and names and proper nouns could also be seen as part of this category) and words that relate to stylistic features (cf. e.g. Scott & Tribble 2006, Scott 2010). Computer generated keywords are also discussed in relation to the 'pre-computer' notion of cultural keywords (cf. e.g. Mahlberg 2007, 2009, Stubbs 2010). Additionally, for the interpretation of keywords the role of the reference corpus as well as the places in the text where the keywords occur can be of importance (cf. Scott 2010, O'Donnell et al. forthcoming).

Studies of computer-generated keywords typically move from the list of keywords to detailed investigations (often by way of concordance analyses) of selected words and to identifying connections between keywords (cf. e.g. Baker 2009). To find groups of keywords, similarities between words are sometimes highlighted with ad hoc categories specific to a particular text (cf. Fischer-Starcke 2009), or keywords can be analysed into categories that are provided by a specific theory, as McEnery (2009) shows with the moral panic discourse model categories. The issue of categorizing keywords can also be addressed by first tagging all the words in the text and the reference corpus with information on the semantic domain they belong to and then calculating keyness by working with the tags (cf. Rayson 2008, Culpeper 2009, Archer et al. 2009).

In the following we will relate the results of a keyword analysis with the results of a key semantic domain analysis. We begin by investigating keywords generated by comparing *CR* against the fictional prose section of the BNC (British National Corpus). We suggest that keywords might be viewed as signals for the building of fictional worlds as well as triggers for thematic concerns of the novel. We then compare the meaning areas that we identified with the help of keywords to key

semantic domains.[2] We do not set out to show that the strength of a keyword analysis lies in finding the most objective classification of words into categories, but we want to argue that, especially for the literary analysis of texts, there is value in discussing where categories usefully take into account subjective interpretations. For the discussion overall, it is important to point out that keywords are a simplified way of looking at meanings in texts. Sinclair (e.g. 2004) stresses that the word is not automatically the most useful unit of meaning, a point also taken up in Stubbs' (2010) discussion of different types of keywords.

3. Keywords in *Casino Royale*

Table 1 contains the top 15 keywords generated by WMatrix comparing *CR* with the fictional prose section of the BNC, a subcorpus of the BNC containing about 15 million words. *CR* contains about 48,000 words. Table 1 shows the frequency of the keyword in *CR*, its frequency in the prose subcorpus of the BNC, and its log-likelihood (LL) value. The table already exemplifies the three very basic groups hinted at above: (i) there are names such as *Bond*, *Le Chiffre*, *Vesper* and *Mathis*, (ii) there are words that give an indication of the 'aboutness' of the text; in Table 1 these are specifically words to do with the casino context, such as *croupier*, *casino*,

Table 1. The top 15 keywords of *CR* compared to the BNC fictional prose section[3]

Rank	Keyword	Freq. CR	Freq. BNC fiction	LL
1	Bond	533	53	5787.70
2	Le_Chiffre	155	0	1786.47
3	Vesper	109	0	1256.29
4	Mathis	105	0	1210.19
5	Croupier	37	4	400.26
6	Casino	38	28	348.17
7	Leiter	30	0	345.77
8	Royale	28	0	322.72
9	Francs	35	57	281.53
10	Smersh	24	0	276.61
11	Banker	34	72	259.31
12	M	38	143	252.85
13	Cards	49	468	243.59
14	He	1276	262488	207.50
15	The	3223	787120	201.83

francs, *cards*; and (iii) there are grammatical words or words that relate to the style of the text, in Table 1 *he* and *the*.

Our analysis aimed to find categories of keywords that are more specific than the three broad groups indicated above and that appear to be useful starting points for the analysis of the text as a whole. We focused on the first 150 keywords, 150 being an arbitrary cut-off.[4] We started by examining the list together with concordance data for the keywords and put words with similar meanings into one group. This approach is similar to the way in which Sinclair (e.g. 2003, 2004) exemplifies the analysis of concordance lines. For each successive keyword the question is whether there is already a group into which it fits or whether a new group is needed. Obviously, in the process, the groups are modified and adjusted to find the best fit for the given set of keywords. Thus with regard to this procedure the cut-off can have an impact on the number of groups. The aim of this method is not the quantification of groups, but the identification of groups.

The groups of keywords are based on meaning relationships in the novel. We suggest that keywords can be broadly grouped in terms of 'fictional world signals' and 'thematic signals'. Table 2 illustrates those two main groups and provides examples of subgroups. 'Thematic signal' keywords are less concrete than 'fictional world' keywords, and as a result are open to wider interpretations. For example, we classified *table* as a 'fictional world' keyword because of its concrete meanings. It is repeatedly used to refer to a gaming table in a casino. In this respect, *table* is a world-building element (akin to some extent to the world-building elements identified by text world theory, cf. Werth 1999, Gavins 2007) and it is difficult to ascribe it any wider textual function. *Gambler*, on the other hand, is an example of a 'thematic signal' keyword. This is because in addition to its concrete sense of the gambler in the casino it can also be interpreted as signalling that the character in question is a risk-taker in life generally. In conjunction with other related keywords, a reader may well determine that one of the themes of the novel is 'taking risks'. (We will discuss this notion of 'gambler' in more detail later on.) The identification of 'fictional world' and 'thematic signal' keywords involves analyzing concordance lines to determine the meanings of the keywords in context. A concordance analysis of *table* shows how it is used to refer to gaming tables, which allows us to categorise it as belonging with all those words which contribute to the construction of the casino context. Whether a word has potential to function as thematic signal is to some extent visible from the concordances but it can also be necessary to take a wider textual context into account, as we will show in Section 4. Likely candidates for words in the thematic signal category tend to be evaluative, metaphorical and potentially ambiguous because of their polysemy.

It will already be clear from this initial explanation that while governed by textual context, the identification of 'fictional world' and 'thematic signal' keywords is

reliant on the interpretative processes of individual readers, which are shaped by both experience of reading (see Carter and Nash 1990) and schematic knowledge (Eysenck and Keane 1990). Indeed, the importance of intuition in combination with quantitative evidence is flagged up by McEnery and Wilson (2001: 11). Interpreting keywords thus involves top-down and bottom-up processing. That is, we need to take into account the local contexts of the keyword in the text, as well as the schematic knowledge it might trigger on the part of the reader and, crucially, what the cumulative effect of this might be within the whole text as a unit of meaning. This may go some way towards explaining why some readers may react differently from others to the novel. To investigate the link between textual cues and cognitive process in more detail, however, studies involving participant testing will be necessary. Thus our suggestions here might be viewed as useful initial categories to be tested in such studies.

After this broad outline of the categories of fictional world and thematic signals we can look for further subgroups. The fictional world category divides into a number of subcategories. So, for example, while keywords such as *cards* and *notes* are clearly objects in the fictional world they are also specific to the casino setting. On this basis, we identified 'casinos and gambling' as subordinate to a higher-order category, 'settings and props', which itself is subordinate to the 'fictional world'

Table 2. Fictional world and thematic signal keywords with examples of subgroups

Category	Example keywords
Fictional world	
Characters:	
Names	*Bond*, *Le Chiffre*, *Vesper*, *Mathis*, *Leiter*, *M*
Body part nouns	*body*, *hand*
Clothes and accessories	*pyjama-coat*, *dinner-jacket*, *benzedrine*
Settings and props:	
Places	*sea*, *Royal-les-Eaux*, *Paris*, *night-club*, *side-road*, *coast*, *room*
Casino and gambling	*gambler*, *casino*, *croupier*, *banco*, *baccarat*, *table*, *chef de partie*, *deuxieme*, *caisse*, *cards*, *notes*
Spies and spying	*gunmen*, *gunman*, *organization*, *S*, *smoke-bomb*, *double*, *bureau*, *agent*, *memorandum*, *trace*
Food and drink	*champagne*, *caviar*
Torture equipment	*cane*, *carpet-beater*, *chair*
Thematic signals	
	gambler, *luck*, *evil*, *black*, *villains*

category. Also subordinate to the 'fictional world' category is the category of 'characters'. Here we included the major characters in the novel while minor characters (e.g. the concierge in the hotel) were treated more like accessories to the plot and subsumed within the constituent groups of the 'settings and props' category. Indeed, major characters too can be seen as inextricably linked with particular categories of setting. *Bond*, for instance, is a major character, but also relates to the specific sub-group of 'spies and spying', as do other characters such as *Le Chiffre* and *Felix Leiter*. This category of 'spies and spying' can be seen to cut across the 'characters' and 'settings and props' superordinate categories, since it also incorporates such keywords as *gunmen* and *smoke-bomb*. Equally, the words that refer to settings and props can relate to different settings. While *table* is often found as a gaming table, it also refers to the other tables, such as the table where Bond and Vesper have dinner together, the small table in the room where Bond is tortured, the table where Bond imagines he will have to face M. The categories are thus by nature fuzzy, which highlights the fact that the creation of meaning in text is not a straightforward process. It also underlines that automatically generated keywords are only based on formal criteria on the textual surface.

We did not investigate grammatical words since we were concerned particularly with lexical words relating to the aboutness of the text (though see, for instance, Culpeper 2009 for a consideration of the importance of grammatical words for both authorial style and text style). There are also words such as *thin* or *directly* that do not seem to fall easily into categories with other keywords. If we were to expand our list of keywords beyond the first 150, we might well find other words that they could group with.

The grouping of keywords relates to a more general issue in corpus linguistics, i.e. the relationship between intuition and preconceived assumption in the analysis of data and the degree to which categories of description are derived in a bottom-up fashion. A point that needs addressing is how 'similar meanings' are defined. The key semantic domains that we will look at in Section 5 below are based on similarities of meanings that are defined through reference to a dictionary. The keyword groups in Table 2 are more subjective in the sense that they have been identified by human analysts (though in accordance with the criteria discussed above). However, the groups reflect the contexts in which the keywords occur and our two broad categories of 'fictional world' keywords and 'thematic signals' indicate different levels of subjectivity. For fictional world keywords that relate to concrete characters and objects the analysis mainly consists of identifying relevant links to characters, objects and places. The thematic category contains mainly abstract and metaphorical meanings that require more complex interpretation. Therefore this category is more reader-centred and relates to the effects that the text creates on the reader at a higher level of abstraction. Our definition

of text-centred, however, does not mean that the keywords can be assigned to groups without reference to the contexts (typically accessed through concordance lines in the first instance) in which they occur. Words that look fairly neutral are, for instance, *coast* or *side-road*, but they turn up in specific scenes to describe the settings, relate to specific people (for example, the word *benzedrine* refers to Le Chiffre's *benzedrine inhaler*[5]), or are part of the props in a scene (e.g. *caviar* when Vesper and Bond have dinner together). Both categories, the text-centred and the reader-centred keywords, are lexically-driven because they are identified on the basis of corpus data. At the same time, they link in with the description of basic elements of novels (characters, settings, themes). Characters and settings are obviously related to the plot of the novel. However, as the keywords are keywords for the whole novel and do not distinguish subsections of the novel, we cannot immediately derive causality or narrative progression from the list, and therefore we do not refer to plot in the analysis that follows.[6]

As shown above, the categories of meanings are fuzzy, which might be seen as an issue with regard to the replicability of the study. However, this fuzziness is important in order to account for the more literary features of the text. Keywords can belong to different categories because of their polysemy. This polysemy in turn is relevant to viewing words as thematic signals, which relates to the fact that polysemy is a concept strongly associated with literariness (Carter and Nash 1990, Carter 2004). It is also possible that keywords have a number of different meanings that contribute to the same category. The word *trace* is an example. Concordance 1 shows all 14 occurrences of *trace*. The patterns in the concordance distinguish three meanings: (i) a verb meaning in lines 11, 12, 13; (ii) the meanings associated with *no trace* illustrated by lines 7, 8, 9, 10 (here we can also include lines 1, 5 and 14, as they are all emphatic expressions to say there is nothing at all), and finally (iii) the expression *a (faint) trace of* that stresses that there is very little or only a hint of something. All three patterns can be seen to contribute to the category 'spies and spying'. The novel is about people trying to find out what others are trying to hide. The most obvious examples are the verb meanings and meanings associated with *no trace*. Some of the examples also seem to relate to characters that are usually strong or expected to keep their cool: only when Bond is tortured does he show *a trace of sluggishness* (line 2), Vesper shows *no trace of self-consciousness* (line 10), when Le Chiffre tortures Bond there is a *trace of impatience* (3) with the man who *showed no trace of emotion* (line 8) in the Casino. In such examples the word *trace* fits two categories: both 'spies and spying' but also the 'character' group. It relates to features that are shared among the characters in this particular setting, but also describes individuals. A particularly clear example is line 4. Here Bond examines his room carefully when he enters it at night and the text tells us that Bond has the qualities that a secret agent needs: "Doing all this, inspecting these

minute burglar-alarms, did not make him feel foolish or self-conscious. He was a secret agent, and still alive thanks to his exact attention to the detail of his profession." However, Bond's exact attention to detail also extends beyond his profession when he observes Vesper's beauty and sees *no trace of make-up* (line 7).

1	hey do at Monte Carlo, there won't be a	**trace** of the mess left in the morning.
2	ly when Bond's tortured spasms showed a	**trace** of sluggishness. He sat for a whil
3	he addressed him again, but now with a	**trace** of impatience. 'that you went
4	ing-desk. Next he examined a faint	**trace** of talcum powder on the inner rim
5	n of our friend from SMERSH. Not a damn	**trace**. He must have got to the villa on
6	'Allez,' said the thin man with a faint	**trace** of impatience. Bond looked him
7	skin was lightly sun-tanned and bore no	**trace** of make-up except on her mouth whi
8	autiously pleased. Le Chiffre showed no	**trace** of emotion. He continued to play l
9	mine the crowd behind him. There was no	**trace** of the gunman, but the huissier wa
10	nts were economical and precise with no	**trace** of self-consciousness. Bond f
11	is voice was a laborious croak. 'Police	**trace** it to you.' Exhausted by the
12	e very long, but it would take hours to	**trace** the ownership to him. And Ves
13	can papers had been at Royale trying to	**trace** the Jamaican millionaire who had d
14	stroy you and the bomb-throwers without	**trace**. Presumably there were other plans

Concordance 1. All 14 occurrences of *trace*

It is essential not to lose sight of the range of meanings of individual keywords in the effort to categorize neatly. Contrary to what is often assumed about corpus linguistics (and levelled at it as a criticism), the point of this exercise was to identify a range of semantic meanings associated with the novel's keywords rather than to quantify these. It underlines that a corpus stylistic analysis should not be solely fixated on quantifying elements of the target text, but should also be sensitive to those aspects of the text which necessitate a more qualitative analysis. Our meaning groups also highlight another point. It is taken for granted by now that corpus data shows how units of meaning are not the same as a word. Extended units of meanings are typically discussed in view of the phraseology of words. At the same time, the fact that words appear together in a text contributes to the meaning that is created by the co-occurrence and the cohesion of the words in the text. Our meaning groups reflect such textual co-occurrence patterns.

4. Thematic signals

The group that we now want to look at in more detail is what we call 'thematic signals'. Here we have included words that on the surface may also be grouped into one of the other groups: *gambler*, *luck* and *odds* fit the gambling context of the casino, *villains* and *evil* could be grouped under spying, and *black* fits into the

group of clothes and accessories, when it refers to Vesper's *black dress*, or Bond's *black satin tie*. However, we singled out these words, because they seem to provide particularly useful clues for interpretation. We will start with the word *luck*.

The noun *luck* occurs 22 times in the novel. It is used in the description of gambling, when people wish each other luck, or specifically to present Bond's view on life. In Chapter 7, the noun occurs eight times altogether. The chapter begins with Bond getting ready for the important poker game and we get a description of how he sees life in relation to luck. We learn that "Bond had always been a gambler" and what he likes about playing and watching others play is that "everything was one's own fault":

(1) *Luck* was a servant and not a master. *Luck* had to be accepted with a shrug or taken advantage of up to the hilt. But it had to be understood and recognized for what it was and not confused with a faulty appreciation of the *odds*, for, at gambling, the deadly sin is to mistake bad play for bad *luck*. And *luck* in all its moods had to be loved and not feared. Bond saw *luck* as a woman, to be softly wooed or brutally ravaged, never pandered to or pursued. (Fleming 1953: 48; our italics)

On the one hand this extract shows how powerful Bond feels in his world described in terms of gambling. On the other hand, the passage also gives indications about his relationship to Vesper. The paragraph continues later with "One day, and he accepted the fact, he would be brought to his knees by love or by luck". This relationship between luck and love is also hinted at earlier in the novel. In Chapter 5 Bond meets Vesper for the first time; the two will have to work together and his feelings about this situation are ambivalent. On the one hand, he is excited about the prospect; on the other he senses the possibility of danger and a threat to his luck: "On an impulse he touched wood".

Bond's view on luck and love is also related to another theme in the novel: the opposition of Bond's human side and his similarity to a dehumanized machine. The link to this theme is visible through the use of the keyword *villains*. This word occurs seven times in the novel and all occurrences are from Chapter 20. The chapter is called "The nature of evil", with *evil* being another keyword that we included in the thematic group. Bond is recovering from his injuries and the character Mathis comes to visit him. Bond has decided to resign from his job and is considering his career. His conversation with Mathis is about his difficulty of dividing the good from the bad and finding where he stands — to Bond it seems "The villains and heroes get all mixed up". Mathis tries to convince Bond that his job is worth doing, in particular because he has now seen such an evil man as Le Chiffre. In this conversation the keywords *villains* and *evil* occur repeatedly, and they also link to the keyword *black*. Mathis refers to the "black targets" that Bond

needs to keep his eye on: "You may want to be certain that the target really is black, but there are plenty of really black targets around". Mathis points out that the black targets become clearer when Bond surrounds himself with human beings, falls in love and has a family, so that he builds up a contrast to the black targets. However, Mathis does not want Bond to become human himself, as he would not be "such a wonderful machine" anymore. The fact that Bond himself is similar to a machine makes it difficult for him to see the grey areas between good and evil. (See also Eco 2003 for a literary critical discussion of the conception of Bond as a machine.)

The theme of the blurring of clear distinctions between good and bad is also made visible in the relationship between Bond and Vesper. Vesper is a double agent. She is a danger to Bond but she is not clearly discernible as a 'black target' but part of the mixing of heroes and villains. At their first meeting Bond seems to sense the danger when he instinctively touches wood. But his qualities as a secret agent, described in the example of *trace* above, fail him in discovering Vesper's involvement with Le Chiffre. At the same time, Bond and Vesper are attracted to one another and they eventually fall in love. The keyword *black* gives some hints that show similarities between Bond and Vesper. When Bond sees Vesper for the first time, she has "a black velvet ribbon on her hat" and wears "shoes of plain black leather"; later in Chapter 15 she wears a "black velvet skirt". Similarly, the description of Bond also refers to black items of dress. In Chapter 8 Bond wears a black satin tie and Vesper a black velvet dress. She jokes about the dress: "if you hear me scream tonight, I shall have sat on a cane chair". This joke alludes to the torture scene to come and takes on even more meaning when it becomes clear that her kidnapping was a trap that resulted in Bond being tortured. Overall, the keywords in the thematic category particularly point to the oppositions between good and bad, human and inhuman, but also to the relationship between love, luck and women within these oppositions.

5. Key semantic domains and text-centred and reader-centred meaning groups

Another approach to identifying meaning groups is the automatic semantic analysis that WMatrix provides. WMatrix assigns a tag to each word specifying the semantic domain the word belongs to. The key comparison is then based on tags so that the results are key semantic domains. Table 3 shows the 20 key semantic domains that result from a comparison of *CR* with the BNC fictional prose section, the same reference corpus as for the keyword comparison above. The symbol […] indicates that only a selection of examples is given; otherwise all words that occur in *CR* and that belong to a given domain are listed. The fact that 'Games' occur as

Table 3. Key semantic domains (critical value of 15.13)

	Semantic domain	Examples
(1)	Games	casino, croupier, gambler, casinos, gamblers, croupiers, ace, juggling, packs_of_cards, vingt-et-un, aces, knock-out, scrabbled, dice
(2)	Grammatical bin	the, and, a, of, […]
(3)	Numbers	two, one, three, five, ten, nine, million, […]
(4)	Money generally	francs, bank, money, bet, cheque, stake, account, twenty_pounds, […]
(5)	Furniture and household fittings	table, chair, bed, tables, seat, desk, drawer, carpet, arm-chair, bath, chairs, curtains, pillow, carpet-beater, curtain, shelves, cane-chair, shutters, cushion, venetian_blinds, […]
(6)	Money and pay	banker, funds, capital, save, investment, saved, paymaster, credit, savings, profit, advance, payment, investments, treasurer, bankers, profits, pay, capitals, takings, pension
(7)	Anatomy and physiology	hand, eyes, face, head, body, mouth, hands, back, arm, hair, arms, feet, heart, teeth, eye, shoulder, […]
(8)	Geographical names	London, Paris, French, Greek, France, Jamaica, […]
(9)	Short and narrow	thin, low, narrow, narrowed, shallow
(10)	Paper documents and writing	cards, card, notes, note, looked_up, letter, memorandum, write, signed, wrote, register, sheaf, […]
(11)	Colour and colour patterns	black, red, white, blue, green, shadow, pink, pale, […]
(12)	Children's games and toys	players, player, kaleidoscope, marionette
(13)	Work and employment: Professionalism	concierge, commissionaire, colleagues, writing-table, colleague, reputation, white-collar, secretary, concierges, bricklayer, doorman
(14)	Smoking and non-medical drugs	cigarette, smoked, cigarettes, tobacco, cigar, smokes, dope, smoking, cigarette-case, Marihuana
(15)	Getting and giving; possession	exchange, exchanged, exchanging, equipped
(16)	Long, tall and wide	thick, long, wide, deep, high, tall, thickly, high-tide, vivacity, thickest, luxuriously
(17)	Chance, luck	luck, chance, chances, happened_to, draws, stake, happen_to
(18)	Success	win, winnings, defeated, winning, came_through, beaten, victory, success, wins, defeat, took_off, make_it, effective, […]
(19)	Parts of buildings	room, door, window, floor, wall, corridor, passage, rooms, doors, windows, entrance_hall
(20)	Linear order	then, first, last, final, third, in_front_of, finally, […]

a key domain means that tags belonging to 'Games' are significantly more frequent in *CR* than in the BNC fictional prose section. As the comparison works on tags, words in a key semantic domain may be both keywords or words that would not appear as key in keyword comparison. In the domain 'Games', *gambler* is a keyword, but *dice* is not (with the same log-likelihood critical value of 15.13, $p < 0.0001$). Additionally, the automatically assigned tags may not always be the most fitting tags in the given context. The word *dice* that occurs just once in *CR* does not appear in the context of games, but in the American English expression *no dice*.

If we compare the automatically generated key semantic domains with our groups of text-centred and reader-centred keywords we can make the following four points:

1. The domain 'Anatomy and physiology' most closely matches one of the keyword groups. It contains the words *body* and *hand* that we included in the fictional world category of 'body part nouns'. In addition to those keywords, the 'Anatomy and physiology' category also collects further body part nouns such as *eyes*, *face*, *head*, *mouth*. The existence of this category is related to the importance of body language in literature (cf. e.g. Korte 1997, Person 1999). However, the fact that this domain comes up as key in comparison with a reference corpus made up of fictional prose indicates that Fleming uses more body part nouns than other authors do in the reference corpus. As the comparison works on tags though, it does not show whether the difference is due to a wider range of body part nouns or the overuse of some specific nouns, which is suggested by the keywords *body* and *hand*. To identify functions of these nouns we need to investigate their contexts in more detail. We will look at these keywords in Section 7 below.
2. There are domains that overlap with the keyword groups and the different groupings of the keywords seem to be to some extent due to the limited number of keywords taken into account in our set of 150. The domain 'Geographical names' overlaps with our category 'Places', but 'Places' include also words other than proper names, such as *room*. Given the set of words we looked at, a separate group for 'Parts of buildings' did not seem to be warranted. Still, 'Parts of buildings' fits with the overall category of 'Settings and props' and is related to 'Places'.
3. There are domains that group words together differently than our keyword groups but also seem to fit less clearly with our keyword groups, because the domains are less specific to the topic of the novel. The domain 'Paper documents and writing' includes *cards* and *notes* relating to the casino context, as well as the word *memorandum* that belongs to the 'Spies and spying' context.[7] Another very broad domain is 'Furniture and household fittings'. It includes

the word *carpet-beater* that we put in the subgroup of 'Torture equipment'. The carpet-beater is not used in its household sense in the novel at all but has a specific function in the torture scene, which cannot be detected by the automatic tagger.

4. Semantic domains that overall seem to fit with the text-centred keywords categories can include some words that are more reader-centred, when these words are used in a non-literal context, e.g. *marionette*. We will look at such examples in more detail in Section 6.

Overall, it seems that the key semantic domains relate most closely to the text-centred categories of keywords, as most domains refer to concrete meanings. The more reader-centred meanings that we find with thematic signals are more abstract and relate most directly to the domain 'Chance, luck'. In the following, we will look at how examples of words that apparently do not fit a category can provide signals for interpretation. We will also look at the two keywords *body* and *hand* in the 'Anatomy and physiology' domain that most clearly matches one of our manually identified groups to search for links with literary critical work on Fleming.

6. Reader-centred words in key domains

We have already mentioned the metaphorical meaning of *gambler*. Related to the meaning of games are also the examples *juggling* and *marionette*. The first is listed in the domain 'Games' that includes mostly casino words, and the second is included in 'Children's games and toys' which also includes the casino words *player/s* together with the word *kaleidoscope*. The words *juggling*, *marionette*, and *kaleidoscope* occur just once each and do not seem to fit the semantic domains in which they are listed. The contexts in which these words appear provide more detail about their meanings. The verb *juggle* in example (2) does not refer to throwing objects in the air but to the way Bond presents the kidnapping in his report to M. In example (3), *marionette* is part of a comparison that describes Bond in the torture scene. Similarly, *kaleidoscope* in example (4) is part of a construction, here the irrealis, that serves to foreground what would otherwise be a fairly conventional narrative trope, thereby increasing the stylistic effectiveness of the narration. While the categorization of *juggling* may be a case where the semantic tagger did not choose the most appropriate of the different meanings of *juggle*, the examples of *marionette* and *kaleidoscope* show how the semantic categories can highlight words with particular functions in the text. Thus key semantic domains may not be the most precise classification of words with regard to the content of the novel,

but they can still point to thematic signals and keywords that require interpretation by the reader.[8]

(2) By *juggling* with the emphasis, he made the kidnapping sound much more Machiavellian than it had been.
(Fleming 1953: 156; our italics)

(3) […] so that Bond screamed and his body jangled in the chair like a *marionette*.
(Fleming 1953: 122; our italics)

(4) He closed his eyes and his thoughts pursued his imagination through a series of carefully constructed scenes as if he was watching the tumbling clips of coloured glass in a *kaleidoscope*.
(Fleming 1953: 56: our italics)

7. Anatomy and physiology (keywords *body* and *hand*)

There are 52 occurrences of *body* in *CR* and, for what is undoubtedly an at times chauvinistic story, our initial hypotheses about character might lead us to expect more references to women's bodies as well as the objectification of these. What is perhaps surprising, then, is that so many of the *body* references (39 of the 52) are to Bond. Seven are references to the body of the heroine, Vesper Lynd, and three refer to the body of the villain, Le Chiffre. (Of the remaining three, two are generic references and one refers to the body of an unnamed villain.) Examining a concordance of those *body* references relating to Bond reveals a number of stylistic effects that potentially contribute towards Bond's characterization. For example, there are ten occurrences of *body* as the head of a noun phrase that forms the subject of a dynamic verb and in each of these cases, the subject noun phrase with *body* as its head could be replaced with the relevant third person pronoun:

```
1                  Bond's whole body writhed and contorted.
2          In the air his whole body turned and with the momentum
3                     his whole body stiffened in a reflex of self-d
4                Then his whole body shrugged
5                  Bond's whole body arched in an involuntary spasm.
6          of his hands and his body sweated with shame.
7                       Then his body sagged and perspiration started
8         Bond screamed and his body jangled in the chair
9    the terrible battering his body had received.
10     made love to him. Bond's body responded, but afterwards
```

Concordance 2. *body* and dynamic verbs

It is stylistically significant that Fleming chooses a subject that is more loaded semantically (this conclusion is predicated on the notion that style arises from choosing one option over another). Eco (2003), in a semiotic analysis of Fleming's Bond novels, picks up on the physicality of Bond, suggesting that "Bond ceases to be a subject for psychiatry and remains at the most a physiological object […], a magnificent machine" (Eco 2003: 35).

Our analysis shows that the preponderance of the noun *body* may be at the root of Eco's observations. The use of *body* rather than the third-person pronoun seems to give a sense that Bond is somehow detached from his physical self. Lines 1, 5, 7 and 8 in Concordance 2 come from the book's infamous torture scene in which Bond's testicles are beaten with a carpet-beater. Bond's body is presented as being somehow separate from his mental self — hence, the implication that an injury to his body is not necessarily an injury to his mind.

1	ain and again so that Bond screamed and	his body	jangled in the chair like a mar
2	was afraid. Afraid that his senses and	his body	would not respond to her sensual
3	ils dug into the palms of his hands and	his body	sweated with shame. Well,
4	f his bathing-trunks and looked down at	his body	. There were only a few traces l
5	vil effects from the terrible battering	his body	had received. He had said that
6	g his eyes, feeling the cold water comb	his body	and his hair. The mirror o
7	d closed his eyes and mentally explored	his body	. The worst pain was in his wris
8	he admitted, why he had waited to give	his body	a chance to respond, why he had
9	eet to his neck, melting the tension in	his body	and calming his still twanging
10	rited and weak in resolve as well as in	his body	. He had had to take too much in
11	and a deep throb of pain started up in	his body	. Mathis realized that he w
12	given a local anaesthetic. The rest of	his body	ached dully as if he had been b
13	d his arms were free. The lower half of	his body	was still shrouded in the oblon
14	less. His buttocks and the underpart of	his body	protruded through the seat of t
15	no colour in Bond's face or anywhere on	his body	above the waist. There was a fa
16	nt, muscles stood out in knots all over	his body	and his toes and fingers clench
17	d perspiration started to bead all oVer	his body	. He uttered a deep groan.
18	nched until they were quite white. Then	his body	sagged and perspiration started
19	rom the passion which had swept through	his body	. He was torn between the desire
20	he wanted coldly to put the repairs to	his body	to the final test. He thought t
21	he was tired. He always knew when	his body	or his mind had had enough and

Concordance 3. All 21 occurrences of *his body*

The foregrounding of Bond's physical self also contributes to a sense of him as a cold and absent character. Concordance 3 contains alls examples of *his body*. Indeed, in those stretches of narration which include references to Bond's internal reflections on his state of health, references to his body are again foregrounded as a result of the way in which Bond's physical self is presented as being separate from his mental self. For example, line 2 in Concordance 3 is free indirect thought demonstrating

that Bond himself makes a distinction between his mental and physical self. Concordance line 21 reiterates this through reference to Bond's beliefs (via the verb form *knew*, of which Bond is the subject). Concordance line 8 is an instance of where *his body* is preferred over the reflexive pronoun *himself*, thereby highlighting the physicality of Bond. The conceptual metaphor underlying line 20 (THE BODY IS A MACHINE; see Lakoff and Johnson 1980, Semino and Swindlehurst 1996) emphasises even further the sense of Bond as somehow inhuman.

If we now consider references to Vesper's body, we can note that these are either from Bond's point of view (lines 1, 6, 7, in which she appears to be objectified by Bond, reflected, for instance, by the words *cold*, *arrogant* and *remote* in line 1), or are narrator-descriptions of Vesper's body which, like references to Bond's body, present her physical self as somehow detached from her mental self (Concordance 4 provides longer stretches of context to make the point clearer):

1 the mirror and wondered about Vesper's morals. He wanted her cold and arrogant **body**. He wanted to see tears and desire in her remote blue eyes and to take the
2 he shutters lit up the room. Only her black hair showed above the sheet and her **body** under the bedclothes was straight and moulded like a stone effigy on a tomb
3 the half-closed shutters and lapped at the secret shadows in the snow of her **body** on the broad bed. Bond awoke in his own room at dawn and for a time he
4 at she could breathe. She was not bound in any other way and she lay quiet, her **body** moving sluggishly with the swaying of the car. Le Chiffre was
5 words she held him more closely to her, murmuring endearments and pressing her **body** down the whole length of his. When he finally rose and bent to smooth
6 knew that she was profoundly, excitingly sensual, but that the conquest of her **body**, because of the central privacy in her, would each time have the sweet tang
7 blue eyes and to take the ropes of her black hair in his hands and bend her long **body** back under his. Bond's eyes narrowed and his face in the mirror looked back

Concordance 4. References to Vesper's body

In concordance line 4, for instance, the noun phrase *her body* is the subject of the non-finite dynamic verb *moving*. The stylistic alternative (not to use *her body* as subject of *moving* but for the subject to be the *she* of the preceding clause) would create the effect of Vesper having some degree of control over her actions, while line 4 stresses the separation of mental and physical self. Concordance line 6 is an instance of Bond regarding Vesper as little more than a physical object, and uses the conceptual metaphor LOVEMAKING IS WAR. It is clear from this that Fleming presents the two main characters as detached, physical objects, as opposed to rounded emotional characters, and this characterization potentially contributes to a conception of the fictional world as a squalid setting that is appropriate for the unpleasant and unfeeling business of espionage.

The other keyword that falls within the key semantic domain of 'Anatomy and Physiology' is *hand*. There are 82 instances of *hand* as a body part (as opposed to *hand* to refer to the cards held by a gambler) and of these, 13 are instances where Bond himself is moving or referring to his hand/s. Five of these are spatial viewpoint related, reflecting Bond's point of view (e.g. *the owner of the hand*, *the hand came out holding*, *there was a hand on his forehead*, *a hand came from behind*, *a broad hairy hand emerging*). Of the fourteen occurrences of *a * hand*, four refer to Bond's hand and contribute to a sense of Bond being once more emotionally detached from his physical self:

1		Bond grunted and moved **a hand**. The thin man slapped him again
2	He put his arms round her and put **a hand** over each breast.	
3	Bond at once relented. He put out **a bandaged hand** and laid it on her knee.	
4		Bond reached out **a steady right hand** and drew the cards

Concordance 5. *a * hand* showing emotional detachment

Nonetheless, there are some instances where Bond himself is moving or referring to his own hand, and where the description is of *his * hand* not *a * hand*. In some cases, using the phrase *a * hand* is semantically marked since the indefinite article suggests that the noun phrase is new information to Bond, which is obviously inappropriate to describe a body part that belongs to him. *His * hand*, therefore, could be perceived as the more conventional form which does not give rise to a conception of Bond's body as distinct from his mental self. This may lead us to conceive of Bond as appearing more complete than in those circumstances when the indefinite noun phrase *a hand* is used. In the novel, there are specific circumstances in which the possessive pronoun precedes the body part (as opposed to an indefinite article). There are 36 instances of *his * hand* in the novel. Thirteen are references to Bond's hand and, of these, four occur in situations where someone other than Bond is doing something to Bond's hand (e.g. *She pressed his hand*) and nine occur in instances where Bond himself is moving or referring to his own hand. Of these nine instances, five occur when Bond is playing cards:

1	fanned the two cards under the curtain of **his hand**.	
2		Bond smiled back and raised **his hand** from the table
3	now seemed to take the colour out of **his hand** as he glanced at the cards.	
4		Bond curled **his right hand** in, glanced briefly
5	his eyes holding Le Chiffre's, reached **his right hand** out a few inches	

Concordance 6. *his * hand* in contexts where Bond is playing cards

Of the remaining four, two refer to Bond holding a gun and two to Bond thinking about his injuries:

1 and ankles and the mark of SMERSH on **his right hand**. He slipped his feet
2 pain was in his wrists and ankles and in **his right hand** where the Russian had
12 full open, the light on and a gun in **his hand**.
24 His last action was to slip **his right hand** under the pillow

Concordance 7. *his * hand* — Bond holding a gun or considering his injuries

The nine instances where Bond himself is doing something to or referring to his own hand show that there is a clearly defined set of circumstances in which *his * hand* is used. These are when Bond is playing cards, when he is holding a gun and when he is considering his injuries. One potential explanation for the use of both *his * hand* and *a * hand* is that the latter is used at times when Bond is in high stress situations, and in his dealings with women (perhaps suggesting a psychological detachment from them).

8. Conclusions

One of the challenges that stylisticians have to face is how to select what is interesting and worthwhile for further study. In this article we have argued that corpus methods can aid the identification of elements of a text worthy of further qualitative analysis, thereby circumventing the problems of selection discussed by Leech and Short (2007). What stylisticians describe as foregrounding effects can be related to quantitative data by comparison with a reference corpus. By using two different approaches, keywords and key semantic domains, we have also worked with different levels of textual detail and considered the interrelationship between the two. The analysis of keywords is based on a more limited set of words than those covered by the semantic domain analysis, but the analysis is more detailed, taking into account more information concerning themes and relationships in the text in question. Key semantic domains have the advantage of grouping together a greater number of words that may not even show in a keyword analysis (even if we had worked with the full keyword list). In this sense the domains also enable the identification of similarities but on a less detailed level. Our distinction into text- and reader-centred keywords also points towards directions for linking corpus analysis with cognitive approaches to text comprehension. We suggest that reader-centred keywords are likely to trigger schematic interpretations at a higher level of abstraction than text-centred keywords. Psycholinguistic experimentation could be used to validate or invalidate this hypothesis.

With regard to *Casino Royale* specifically, our analysis provides some measure of objective support for comment on Fleming's work by literary critics. For example, the keyness of the 'Anatomy and physiology' domain supports Eco's (2003) notion of Bond as a machine. Beyond this, our corpus analysis uncovers some of

the triggers of characterization in the novel, and begins to account for readers' conceptions of the fictional world and potential reactions to it. Finally, we hope to have shown that a corpus stylistic approach provides a useful means of relatively objectively determining starting-points for a qualitative analysis. We suggest that this makes it a powerful complementary methodology for more traditional methods of literary criticism.

Notes

1. Louw attributes the origins of the term 'semantic prosody' to John Sinclair.

2. To obtain an electronic copy of the text, the book was scanned manually. This work was supported by a grant from the University of Huddersfield's Selective Assistance to Staff fund, for which we are grateful. We would also like to thank Catherine Smith for help with the tokenization of the text, and Paul Rayson for his support while using WMatrix.

3. Note that WMatrix identifies multiword units (MWUs), so that *Le Chiffre* is identified as one unit. MWUs also affect the counts of other keywords, such as *Leiter*, because the sequence *Felix Leiter* is identified as a MWU, and therefore not included in the counts for the surname alone.

4. For the first 150 keywords the LL value is ≥ 31.31.

5. We included the word *benzedrine* in the category 'clothes and accessories' because in all three occurrences it refers to the inhaler that Le Chiffre carries.

6. For the use of keywords for the analysis of narrative progression see Toolan (2009).

7. WMatrix has an option that allows one to create a personal lexicon so that words can be assigned to different semantic categories; for our study we worked with the standard settings. Further studies might usefully investigate the options of modifying the lexicon. The groups that we suggest here could be a starting point for such a modification.

8. Cf. Hardie et al. (2007) on the use of semantic tagging and WMatrix for the identification of possible source and target domains in the analysis of metaphors. Hardie et al. (2007) treat automatically generated semantic fields as roughly equivalent to the domains of conceptual metaphor theory (on which, see Lakoff and Johnson 1980).

References

Adolphs, Svenja and Ron Carter. 2002. Corpus stylistics: point of view and semantic prosodies in *To The Lighthouse*. *Poetica* 58: 7–20.

Archer, Dawn, Jonathan Culpeper and Paul Rayson. 2009. Love — 'a familiar or a devil'? An exploration of key domains in Shakespeare's Comedies and Tragedies. In *What's in a wordlist? Investigating word frequency and keyword extraction*, Dawn Archer (ed.). Farnham: Ashgate, 137–157.

Baker, Paul. 2009. The question is, how cruel is it? Keywords, fox hunting and the House of Commons. In *What's in a word-list? Investigating word frequency and keyword extraction*, Dawn Archer (ed.). Farnham: Ashgate, 125–136.

Biber, Douglas. 2011. Corpus linguistics and the scientific study of literature: back to the future? *Scientific Study of Literature* 1 (1): 15–23.

Carter, Ron. 2004. *Language and Creativity: The Art of Common Talk*. London: Routledge.

Carter, Ron and Walter Nash. 1990. *Seeing through Language: A Guide to Styles of English Writing*. Oxford: Blackwell.

Culpeper, Jonathan. 2009. Keyness: words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*, *International Journal of Corpus Linguistics* 14 (1): 29–59.

Eco, Umberto. 2003. Narrative structure in Ian Fleming. In *The James Bond Phenomenon: A Critical Reader*, Christopher Lindner (ed.). Manchester: Manchester University Press, 34–55.

Eysenck, Michael W. and Mark T. Keane. 1990. *Cognitive Psychology: A Student's Handbook*. Hove and London: Lawrence Erlbaum Associates.

Fischer-Starcke, Bettina. 2009. Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*: A corpus-stylistic analysis, *International Journal of Corpus Linguistics* 14(4): 492–523.

Freeman, Donald C. 1970. Linguistic approaches to literature. In *Linguistics and Literary Style*, Donald C. Freeman, (ed.). New York: Holt, Reinhart and Winston, 3–17.

Gavins, Joanna. 2007. *Text World Theory: An Introduction*. Edinburgh: Edinburgh University Press.

Hardie, Andrew, Veronika Koller, Paul Rayson and Elena Semino. 2007. Exploiting a semantic annotation tool for metaphor analysis. In *Proceedings of the Corpus Linguistics 2007 Conference*, Matthew Davies, Paul Rayson, Susan Hunston and Pernilla Danielsson. (eds.). Birmingham: University of Birmingham. [Available at: http://corpus.bham.ac.uk/corpling-proceedings07/paper/49_Paper.pdf accessed July 2010]

Hori, Masahiro. 2004. *Investigating Dickens' Style. A Collocational Analysis.* Basingstoke: Palgrave Macmillan.

Korte, Barbara. 1997. *Body Language in Literature*. Toronto: University of Toronto Press.

Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: Chicago University Press.

Leech, Geoffrey and Michael Short. 2007. *Style in Fiction: A Linguistic Introduction to English Fictional Prose*, 2nd edn. London: Pearson Education.

Louw, Bill. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In *Text and Technology: In Honour of John Sinclair*, Mona Baker, Gill Francis and Elena Tognini-Bonelli. (eds.). Amsterdam: John Benjamins, 155–176.

Mahlberg, Michaela. 2007. Lexical items in discourse: identifying local textual functions of *sustainable development*. In *Text, Discourse and Corpora. Theory and Analysis*, Hoey, M., Mahlberg, M., Stubbs, M., Teubert, W. London: Continuum, 191–218.

Mahlberg, Michaela. 2009. Patterns in news stories: a corpus approach to teaching discourse analysis. In *Using Corpora to Learn about Language and Discourse*, Linda Lombardo (ed.). Bern: Peter Lang, 99–132.

Mahlberg, Michaela. forthcoming. Corpus analysis of literary texts. In *The Encyclopedia of Applied Linguistics*, Carole A. Chapelle (ed.). Oxford: Wiley-Blackwell.

McEnery, Tony. 2009. Keywords and moral panics: Mary Whitehouse and media censorship. In *What's in a word-list? Investigating word frequency and keyword extraction*, Dawn Archer (ed.). Farnham: Ashgate, 93–124.

McEnery, Tony and Andrew Wilson. 2001. *Corpus Linguistics*, 2nd edn. Edinburgh: Edinburgh University Press.

McIntyre, Dan. forthcoming. Corpora and literature. In *The Encyclopedia of Applied Linguistics*, Carole A. Chapelle (ed.). Oxford: Wiley-Blackwell.

McIntyre, Dan and Brian Walker. 2010. How can corpora be used to explore the language of poetry and drama? In *The Routledge Handbook of Corpus Linguistics*, Michael McCarthy and Anne O'Keefe (eds.). Abingdon: Routledge, 516–530.

Montoro, Rocío. 2007. The stylistics of cappuccino fiction: a socio-cognitive perspective. In *Contemporary Stylistics*, Marina Lambrou and Peter Stockwell (eds.). London: Continuum, 68–80.

Nash, Walter. 1990. *Language in Popular Fiction*. London: Routledge.

O'Donnell, Matthew Brook, Mike Scott, Michaela Mahlberg and Michael Hoey. forthcoming. Exploring text-initial words, clusters and concgrams in a Newspaper Corpus. *Corpus Linguistics and Linguistic Theory*.

O'Halloran, Kieran. 2007. The subconscious in James Joyce's "Eveline": a corpus stylistic analysis that chews on the "Fish hook". *Language and Literature* 16 (3): 227–244.

Person, Raymond F. Jr. (1999) *Structure and Meaning in Conversation and Literature*. Lanham: University Press of America.

Rayson, Paul. 2008. From key words to key semantic domains. *International Journal of Corpus Linguistics* 13 (4): 519–549.

Scott, Mike. 2010. Problems in investigating keyness, or clearing the undergrowth and marking out trails. In *Keyness in Texts*, Marina Bondi and Mike Scott. (eds.). Amsterdam: John Benjamins, 43–57.

Scott, Mike and Christopher Tribble. 2006. *Textual Patterns. Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.

Semino, Elena and Mick Short. 2004. *Corpus Stylistics. Speech, Writing and Thought Presentation in a Corpus of English Writing*. London: Routledge.

Semino, Elena and Kate Swindlehurst. 1996. Metaphor and Mind Style in Ken Kesey's *One Flew Over the Cuckoo's Nest*. *Style* 30 (1): 143–166.

Sinclair, John. 2003. *Reading Concordances. An Introduction.* Harlow: Pearson Education.

Sinclair, John. 2004. *Trust the Text. Language, Corpus and Discourse.* London: Routledge.

Stubbs, Michael. 2005. Conrad in the computer: examples of quantitative stylistic methods. *Language and Literature* 14 (1): 5–24.

Stubbs, Michael. 2010. Three concepts of keywords. In *Keyness in Texts*, Marina Bondi and Mike Scott. (eds.). Amsterdam: John Benjamins, 21–42.

Toolan, Michael. 2009. *Narrative Progression in the Short Story: A Corpus Stylistic Approach*. Amsterdam: John Benjamins.

Walker, Brian. 2010. WMatrix, key concepts and the narrators in Julian Barnes's *Talking It Over*. In *Language and Style*, Dan McIntyre and Beatrix Busse. (eds.). Basingstoke: Palgrave, 364–387.

Werth, Paul. 1999. *Text Worlds: Representing Conceptual Space in Discourse*. London: Longman.

Author's addresses

Michaela Mahlberg
School of English Studies
University of Nottingham
University Park
Nottingham
NG7 2RD

michaela.mahlberg@nottingham.ac.uk

Dan McIntyre
School of Music, Humanities and Media
University of Huddersfield
Queensgate
Huddersfield
HD1 3DH

d.mcintyre@hud.ac.uk